

**МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ**  
**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ**  
**УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**  
**«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»**  
**(РУТ (МИИТ))**



Рабочая программа дисциплины (модуля),  
как компонент образовательной программы  
базового высшего образования  
по направлению подготовки  
09.03.01 Информатика и вычислительная техника,  
утвержденной первым проректором ФГАОУ ВО  
РУТ (МИИТ) Тимониным В.С.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

**Технологии хранения больших данных**

Направление подготовки: 09.03.01 Информатика и вычислительная техника

Направленность (профиль): IT-сервисы и технологии обработки данных на транспорте (Российско-Китайская программа)

Форма обучения: Очная

Рабочая программа дисциплины (модуля) в виде  
электронного документа выгружена из единой  
корпоративной информационной системы управления  
университетом и соответствует оригиналу

Простая электронная подпись, выданная ФГАОУ ВО РУТ  
(МИИТ)

ID подписи: 35225

Подписал: заместитель директора Игольников Борис  
Владимирович

Дата: 16.10.2026

## 1. Общие сведения о дисциплине (модуле).

Целями освоения учебной дисциплины являются формирование у обучающихся системных знаний и практических навыков в области технологий хранения и обработки больших данных для проведения аналитических исследований в транспортной отрасли. В рамках курса рассматриваются современные архитектуры хранения данных, распределенные файловые системы, NoSQL-базы данных, озера данных (Data Lake) и хранилища данных (Data Warehouse), методы статистического анализа данных, а также критерии выбора алгоритмов для различных задач анализа транспортных данных. Особое внимание уделяется работе с большими объемами транспортных телематических данных, логистических потоков и данных GPS/ГЛОНАСС, а также практическим аспектам применения технологий больших данных для решения прикладных задач отрасли.

Задачами освоения дисциплины являются:

Формирование теоретических знаний об архитектурах хранения больших данных, распределенных системах и моделях данных.

Освоение технологий и инструментов для работы с большими данными (Hadoop, Spark, Hive, HBase, Cassandra, MongoDB, ClickHouse).

Развитие практических навыков построения хранилищ данных и озер данных для транспортных задач.

Формирование компетенций по применению методов статистического анализа данных на больших объемах информации.

Освоение критериев выбора алгоритмов анализа данных в зависимости от типа решаемой транспортной задачи и характеристик данных.

Приобретение навыков проведения аналитических исследований с применением технологий больших данных в транспортной отрасли.

Дисциплина предназначена для формирования компетенций, необходимых для эффективного проведения аналитических исследований с применением технологий больших данных в транспортных организациях.

## 2. Планируемые результаты обучения по дисциплине (модулю).

Перечень формируемых результатов освоения образовательной программы (компетенций) в результате обучения по дисциплине (модулю):

**ПК-6** - Способен проводить аналитические исследования с применением технологий больших данных.

Обучение по дисциплине (модулю) предполагает, что по его результатам обучающийся будет:

**Знать:**

Методы статистического анализа данных, применяемые для обработки больших объемов транспортной информации.

Критерии выбора алгоритмов для различных задач анализа больших данных на транспорте.

**Уметь:**

Применять методы статистического анализа данных для исследования транспортных потоков, телематических данных и логистических процессов.

Выбирать оптимальные алгоритмы и технологии хранения данных в зависимости от типа решаемой задачи и характеристик транспортных данных.

**Владеть:**

Навыками работы с распределенными системами хранения и обработки данных (Hadoop, Spark) и NoSQL-базами данных.

Техниками построения хранилищ данных и озер данных для транспортных организаций

**3. Объем дисциплины (модуля).**

**3.1. Общая трудоемкость дисциплины (модуля).**

Общая трудоемкость дисциплины (модуля) составляет 4 з.е. (144 академических часа(ов)).

**3.2. Объем дисциплины (модуля) в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:**

| Тип учебных занятий                                       | Количество часов |            |
|-----------------------------------------------------------|------------------|------------|
|                                                           | Всего            | Семестр №3 |
| Контактная работа при проведении учебных занятий (всего): | 64               | 64         |
| В том числе:                                              |                  |            |
| Занятия лекционного типа                                  | 16               | 16         |
| Занятия семинарского типа                                 | 48               | 48         |

**3.3. Объем дисциплины (модуля) в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации**

образовательной программы на иных условиях, при проведении промежуточной аттестации составляет 80 академических часа (ов).

3.4. При обучении по индивидуальному учебному плану, в том числе при ускоренном обучении, объем дисциплины (модуля) может быть реализован полностью в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации.

#### 4. Содержание дисциплины (модуля).

##### 4.1. Занятия лекционного типа.

| №<br>п/п | Тематика лекционных занятий / краткое содержание                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1        | Тема 1. Введение в технологии хранения больших данных.<br>Рассматриваемые вопросы:<br>- Понятие больших данных (Big Data) и их характеристики (Volume, Velocity, Variety, Veracity, Value).<br>- Особенности транспортных больших данных (GPS/ГЛОНАСС, логистика, телематика, биллинг).<br>- Эволюция систем хранения данных: от файловых систем к распределенным.<br>- Классификация систем управления базами данных.<br>- Современные тренды и вызовы в хранении больших данных на транспорте. |
| 2        | Тема 2. Жизненный цикл данных в транспортных системах.<br>Рассматриваемые вопросы:<br>- Этапы жизненного цикла данных: сбор, хранение, обработка, анализ, архивация.<br>- Особенности каждого этапа для транспортных данных.<br>- Политики управления данными (Data Governance).<br>- Роли и ответственность в управлении данными.                                                                                                                                                               |
| 3        | Тема 3. Архитектуры хранения больших данных: Data Warehouse.<br>Рассматриваемые вопросы:<br>- Концепция хранилища данных (Data Warehouse).<br>- Схемы звезда и снежинка.<br>- Преимущества и ограничения Data Warehouse.<br>- Применение Data Warehouse для транспортной аналитики.<br>- Примеры реализаций на транспорте.                                                                                                                                                                       |
| 4        | Тема 4. Архитектуры хранения больших данных: Data Lake.<br>Рассматриваемые вопросы:<br>- Концепция озера данных (Data Lake).<br>- Преимущества и ограничения Data Lake.<br>- Сравнение Data Lake и Data Warehouse.<br>- Применение Data Lake для хранения сырых транспортных данных.<br>- Кейсы использования на транспорте.                                                                                                                                                                     |
| 5        | Тема 5. Архитектуры хранения больших данных: Data Lakehouse.<br>Рассматриваемые вопросы:<br>- Концепция Data Lakehouse: объединение Data Lake и Data Warehouse.                                                                                                                                                                                                                                                                                                                                  |

| №<br>п/п | Тематика лекционных занятий / краткое содержание                                                                                                                                                                                                                                                                                                                                                                                                            |
|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|          | <ul style="list-style-type: none"> <li>- Преимущества и ограничения.</li> <li>- Технологии реализации Lakehouse (Delta Lake, Iceberg, Hudi).</li> <li>- Применение для транспортных задач.</li> <li>- Перспективы развития Lakehouse на транспорте.</li> </ul>                                                                                                                                                                                              |
| 6        | <p>Тема 6. Сравнительный анализ архитектур хранения для транспортных задач.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Критерии выбора архитектуры хранения.</li> <li>- Сравнительный анализ Data Warehouse, Data Lake, Lakehouse.</li> <li>- Выбор архитектуры в зависимости от типа транспортной задачи.</li> <li>- Гибридные подходы и их применение.</li> <li>- Рекомендации для транспортных организаций.</li> </ul> |
| 7        | <p>Тема 7. Распределенные файловые системы: HDFS.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Понятие распределенной файловой системы.</li> <li>- HDFS: архитектура и принципы работы (NameNode, DataNode).</li> <li>- Репликация и отказоустойчивость в HDFS.</li> <li>- Работа с большими файлами в HDFS.</li> <li>- Интеграция HDFS с инструментами обработки.</li> </ul>                                               |
| 8        | <p>Тема 8. Экосистема Hadoop.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Обзор экосистемы Hadoop.</li> <li>- YARN: управление ресурсами и планирование задач.</li> <li>- Hive: SQL-интерфейс для Hadoop.</li> <li>- HBase: NoSQL-база данных на Hadoop.</li> <li>- Применение экосистемы Hadoop для транспортных данных.</li> </ul>                                                                                       |

## 4.2. Занятия семинарского типа.

### Практические занятия

| №<br>п/п | Тематика практических занятий/краткое содержание                                                                                                                                                                                                                                                                                                                                                              |
|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1        | <p>Тема 1. Установка и настройка среды для работы с большими данными.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Установка Docker и Docker Compose.</li> <li>- Настройка окружения для Hadoop, Spark, ClickHouse.</li> <li>- Создание кластера в контейнерах.</li> <li>- Проверка работоспособности всех компонентов.</li> <li>- Первые шаги в командной строке.</li> </ul> |
| 2        | <p>Тема 2. Введение в экосистему Hadoop.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Знакомство с архитектурой HDFS.</li> <li>- Базовые команды работы с HDFS (ls, put, get, cat, mkdir).</li> <li>- Загрузка и выгрузка данных в HDFS.</li> <li>- Работа с правами доступа в HDFS.</li> <li>- Мониторинг HDFS.</li> </ul>                                                   |
| 3        | <p>Тема 3. Продвинутая работа с HDFS.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Репликация файлов и управление репликацией.</li> <li>- Работа с большими файлами (блоки, партиции).</li> </ul>                                                                                                                                                                             |

| №<br>п/п | Тематика практических занятий/краткое содержание                                                                                                                                                                                                                                                                                                                                                                             |
|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|          | <ul style="list-style-type: none"> <li>- Сжатие данных в HDFS (gzip, snappy, lzo).</li> <li>- Настройка квот и пространства имен.</li> <li>- Мониторинг состояния кластера.</li> </ul>                                                                                                                                                                                                                                       |
| 4        | <p>Тема 4. Введение в Apache Hive.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Установка и настройка Hive.</li> <li>- Создание таблиц в Hive (внутренние и внешние).</li> <li>- Загрузка данных в Hive из HDFS.</li> <li>- Выполнение SQL-запросов в Hive (SELECT, GROUP BY, JOIN).</li> <li>- Практика: загрузка транспортного датасета.</li> </ul>                                        |
| 5        | <p>Тема 5. Продвинутая работа с Apache Hive.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Секционирование таблиц в Hive.</li> <li>- Использование ведра (bucketing) для оптимизации.</li> <li>- Сложные запросы с подзапросами и CTE.</li> <li>- Интеграция Hive с другими инструментами.</li> <li>- Оптимизация производительности запросов.</li> </ul>                                     |
| 6        | <p>Тема 6. Введение в Apache HBase.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Архитектура HBase (Region Server, HMaster, ZooKeeper).</li> <li>- Создание таблиц в HBase.</li> <li>- Вставка и извлечение данных (put, get, scan).</li> <li>- Работа с фильтрами и сканированием.</li> <li>- Практика: хранение телематических данных.</li> </ul>                                          |
| 7        | <p>Тема 7. Интеграция HBase с Hive.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Создание внешних таблиц в Hive над HBase.</li> <li>- Выполнение SQL-запросов к данным в HBase.</li> <li>- Сравнение производительности Hive и HBase.</li> <li>- Сценарии использования HBase на транспорте.</li> <li>- Практика: гибридный запрос.</li> </ul>                                               |
| 8        | <p>Тема 8. Введение в MongoDB: установка и базовые операции.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Установка и настройка MongoDB.</li> <li>- Структура документов в MongoDB.</li> <li>- CRUD-операции в MongoDB (insert, find, update, delete).</li> <li>- Создание индексов для оптимизации запросов.</li> <li>- Практика: загрузка транспортных данных.</li> </ul>                  |
| 9        | <p>Тема 9. Агрегационные запросы в MongoDB.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Конвейеры агрегации (aggregation pipeline).</li> <li>- Операторы агрегации: match, match, group, project, project, sort, \$limit.</li> <li>- Сложные агрегации с несколькими этапами.</li> <li>- Анализ транспортных данных с агрегациями.</li> <li>- Практика: анализ пассажиропотоков.</li> </ul> |
| 10       | <p>Тема 10. Оптимизация MongoDB для транспортных данных.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Настройка производительности MongoDB.</li> <li>- Использование составных индексов.</li> <li>- Оптимизация запросов с explain().</li> </ul>                                                                                                                                             |

| №<br>п/п | Тематика практических занятий/краткое содержание                                                                                                                                                                                                                                                                                                                                                                                             |
|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|          | <ul style="list-style-type: none"> <li>- Шардирование и репликация в MongoDB.</li> <li>- Практика: оптимизация запросов.</li> </ul>                                                                                                                                                                                                                                                                                                          |
| 11       | <p>Тема 11. Введение в ClickHouse: установка и базовые операции.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Установка и настройка ClickHouse.</li> <li>- Создание таблиц с различными движками (MergeTree, Log).</li> <li>- Загрузка данных в ClickHouse.</li> <li>- Выполнение базовых запросов (SELECT, WHERE, ORDER BY).</li> <li>- Практика: загрузка транспортного датасета.</li> </ul>               |
| 12       | <p>Тема 12. Продвинутая работа с ClickHouse.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Создание таблиц с партиционированием и сортировкой.</li> <li>- Использование материализованных представлений.</li> <li>- Работа с временными рядами в ClickHouse.</li> <li>- Оптимизация запросов с EXPLAIN.</li> <li>- Практика: сложные аналитические запросы.</li> </ul>                                        |
| 13       | <p>Тема 13. Аналитика транспортных данных в ClickHouse.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Агрегация больших объемов данных.</li> <li>- Группировка по времени и категориям.</li> <li>- Вычисление скользящих средних и трендов.</li> <li>- Анализ пассажиропотоков и загрузки.</li> <li>- Визуализация результатов аналитики.</li> </ul>                                                          |
| 14       | <p>Тема 14. Сравнительный анализ NoSQL-решений для транспортных данных.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Сравнение MongoDB, ClickHouse, Cassandra.</li> <li>- Критерии выбора для различных задач.</li> <li>- Тестирование производительности на транспортных данных.</li> <li>- Выбор решения для конкретной транспортной задачи.</li> <li>- Документирование результатов сравнения.</li> </ul> |
| 15       | <p>Тема 15. Введение в Apache Cassandra.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Архитектура Cassandra: принципы распределения, репликации, согласованности.</li> <li>- Установка и настройка Cassandra.</li> <li>- Создание keyspace и таблиц.</li> <li>- CRUD-операции в Cassandra.</li> <li>- Практика: работа с транспортными данными.</li> </ul>                                                   |
| 16       | <p>Тема 16. Продвинутая работа с Cassandra.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Моделирование данных для Cassandra (партиционирование, кластеризация).</li> <li>- Оптимизация запросов в Cassandra.</li> <li>- Настройка согласованности и доступности.</li> <li>- Масштабирование Cassandra-кластера.</li> <li>- Практика: оптимизация запросов на транспортных данных.</li> </ul>                 |
| 17       | <p>Тема 17. Введение в Apache Spark.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> <li>- Установка и настройка Spark.</li> <li>- Архитектура Spark: драйвер, исполнители.</li> <li>- Spark Session и Spark Context.</li> </ul>                                                                                                                                                                                      |

| №<br>п/п | Тематика практических занятий/краткое содержание                                                               |
|----------|----------------------------------------------------------------------------------------------------------------|
|          | - Основные операции с DataFrames (select, filter, groupBy).<br>- Практика: загрузка и первичный анализ данных. |

#### 4.3. Самостоятельная работа обучающихся.

| №<br>п/п | Вид самостоятельной работы                                                            |
|----------|---------------------------------------------------------------------------------------|
| 1        | Изучение литературы и документации по Hadoop, Spark, Hive, HBase.                     |
| 2        | Работа с официальной документацией NoSQL-баз данных (MongoDB, Cassandra, ClickHouse). |
| 3        | Изучение методов статистического анализа больших данных.                              |
| 4        | Выполнение домашних заданий по работе с распределенными системами хранения            |
| 5        | Подготовка презентации по результатам проекта для защиты.                             |
| 6        | Подготовка к промежуточной аттестации.                                                |
| 7        | Подготовка к промежуточной аттестации.                                                |
| 8        | Подготовка к текущему контролю.                                                       |

#### 4.4. Примерный перечень тем курсовых работ

Проектирование и реализация озера данных (Data Lake) для хранения GPS/ГЛОНАСС-данных транспортного предприятия.

Сравнительный анализ производительности ClickHouse и MongoDB для аналитики пассажиропотоков.

Разработка ETL-пайплайна для загрузки и трансформации транспортных телематических данных с использованием Apache Airflow и Spark.

Кластерный анализ маршрутов общественного транспорта на основе данных GPS с использованием Spark MLlib.

Построение хранилища данных (Data Warehouse) для агрегирования и анализа логистических показателей транспортной компании.

Разработка системы мониторинга качества данных в озере данных транспортной организации с использованием Great Expectations.

Прогнозирование пассажиропотоков городского транспорта с использованием Spark ML и статистических методов.

Обнаружение аномалий в данных телеметрии подвижного состава с использованием Isolation Forest в Spark.

Проектирование архитектуры Lakehouse для хранения и анализа мультимодальных транспортных данных с использованием Delta Lake.

Разработка интерактивного дашборда для мониторинга транспортных потоков на основе данных из ClickHouse и Apache Superset.

5. Перечень изданий, которые рекомендуется использовать при освоении дисциплины (модуля).

| № п/п | Библиографическое описание                                                                                                                                                                                                                          | Место доступа                                                                     |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| 1     | Исабекова, О. А. Технологии формирования межотраслевого и межгосударственного единого информационного пространства : учебно-методическое пособие / О. А. Исабекова, Е. М. Дворникова. — Москва : РТУ МИРЭА, 2024. — 68 с. — ISBN 978-5-7339-2375-8. | <a href="https://e.lanbook.com/book/464624">https://e.lanbook.com/book/464624</a> |
| 2     | Системы искусственного интеллекта и большие данные. Раздел «Большие данные» : учебное пособие / А. А. Парамонов, И. А. Юрченков, Б. А. Крынецкий, И. В. Есипов. — Москва : РТУ МИРЭА, 2025. — 147 с. — ISBN 978-5-7339-2499-1.                      | <a href="https://e.lanbook.com/book/493550">https://e.lanbook.com/book/493550</a> |

6. Перечень современных профессиональных баз данных и информационных справочных систем, которые могут использоваться при освоении дисциплины (модуля).

<https://habr.com/ru> - база знаний в виде статей, обзоров

<http://library.miit.ru/> - электронно-библиотечная система научно-технической библиотеки МИИТ

<https://superset.apache.org/user-docs/intro/#documentation> – документация Apache Superset

<https://www.postgresql.org/> - официальная страница PostgreSQL

<https://yandex.cloud/ru/docs/datalens/> - официальная документация Yandex Datalens

<https://docs.docker.com/> - официальная документация Docker  
<https://www.metabase.com/docs/latest/> - официальная документация Metabase  
<https://tableau.pro/> - официальный сайт Tableau с лучшими практиками проектирования дашбордов  
<https://revealthedata.com/blog/> - Блог о визуализации данных и BI-системах

7. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, необходимого для освоения дисциплины (модуля).

MS Office (Excel, Word, PowerPoint)  
Apache Superset  
Yandex Datalens  
Metabase  
PostgreSQL  
Docker  
Apache Airflow  
Браузер Chrome / Яндекс браузер  
Draw.io

8. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю).

Компьютер преподавателя с доступом к сети Интернет  
Компьютеры студентов с установленным Python-окружением и табличным процессором  
Доступ к учебной базе событийных данных и экспериментальным выборкам  
Доступ к PostgreSQL и BI-системе для визуализации результатов  
Доступ к облачной виртуальной машине или Docker-контейнерам для воспроизводимого анализа  
Проектор и экран для проектора  
Маркерная доска

9. Форма промежуточной аттестации:

Зачет в 3 семестре.

10. Оценочные материалы.

Оценочные материалы, применяемые при проведении промежуточной аттестации, разрабатываются в соответствии с локальным нормативным актом РУТ (МИИТ).

Авторы:

руководитель образовательной  
программы

О.Б. Проневич

Согласовано:

Директор

Д.В. Паринов

Заместитель директора

Б.В. Игольников

Председатель учебно-методической  
комиссии

Д.В. Паринов